DOCUMENT RESUME

ED 261 331 CG 018 530

AUTHOR Silverman, William H.; And Others
TITLE Influence of Assessment Center Methods on Assessor Ratings.
PUB DATE Aug 85
NOTE 18p.; Paper presented at the Annual Convention of the American Psychological Association (93rd, Los Angeles, CA, August 23-27, 1985).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Assessment Centers (Personnel); *Cognitive Processes; *Evaluation Methods; *Evaluators; Personnel Selection

ABSTRACT

One potential source of variation in assessment center ratings may be the way the assessor organizes and processes information. To examine how assessment center methods affect the way assessors organize and process assessment center information and the ratings they make, independent groups of assessors (N=24) used one of two models for integrating assessment center information into a final evaluation of candidates (N=90) for a retail organization's department manager position. Half of the candidates were assessed using the within exercise model in which candidates were rated on each dimension upon completion of each exercise. The other half were assessed using the modified American Telephone and Telegraph Company model in which assessors shared their observations and rated candidates at the completion of all exercises. Each candidate was rated by four assessors who observed candidates in three exercises (In-Basket and two role plays) and assessed them on six dimensions (problem analysis, problem solution, leadership, people sensitivity, sense of urgency, and communication). Data indicated differences in the convergent and discriminant validities, and differences in the factor structures of the ratings for the two models. The pattern of results suggests that the two assessment center models for integrating information affected both the way the assessors organized the assessment center information and the obtained ratings. (NRB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED261331

Influence of Assessment Center Methods on Assessor Ratings

William H. Silverman
Giant Food, Inc.
Landover, MD

Anthony Dalessio
Old Dominion University

Steven B. Woods
Old Dominion University

Rudolph L. Johnson, Jr.
Giant Food, Inc.
Landover, MD

Running head: ASSESSMENT CENTER METHODS

CE 018530

Paper presented at the Annual Meeting of the American Psychological Association, Los Angeles, August, 1985

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Anthony Dalessio

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Influence of Assessment Center Methods

on Assessor Ratings

Abstract

The purpose of the present study was to examine how assessment center methods affect the way assessors organize and process assessment center information and affect the ratings they make. Independent groups of assessors underwent one of two models for integrating assessment center information into a final evaluation of the candidate. Data indicated differences in the convergent and discriminant validities, and differences in the factor structures of the ratings for the two models. The pattern of results suggested that the two assessment center models for integrating information affected the way the assessors organized the assessment center information and affected the obtained ratings. Future research should consider assessment center methods and assessors as a source of variation in assessment center ratings.

Running head: ASSESSMENT CENTER METHODS

Influence of Assessment Center Methods on Assessor Ratings

In recent years a number of authors have argued persuasively that performance ratings are influenced to a large extent by the way the rater selects, organizes, stores, and recalls information (Feldman, 1981; Landy & Farr, 1980; Rush, Thomas, & Lord, 1977). To the extent that this is indeed the case, ratings may tell us as much about the cognitive processes of the rater as they do about the ability and performance of the ratee. The cognitive processes of the rater may be influenced by a wide variety of factors related to the stimulus, the perceiver, and the situation (Shack, 1983).

Although the cognitive processes of the rater have been considered in the job performance evaluation literature, this issue has not been considered in the assessment center. A potential source of variation in assessment center ratings may be the way the assessor selects, organizes, stores, and recalls information. These cognitive processes of the assessor may be influenced to a large extent by assessment center methods and procedures. Despite the differences in methodologies which exist among assessment centers, little concern has been shown for the potential impact that these variations may have on the way the assessor processes information and the ratings subsequently obtained. Consequently the purpose of the present study was to examine one aspect of assessment center methodology and its impact on assessor ratings.

One of the most important aspects of the assessment center

is the way in which information obtained by the assessors is integrated into the final overall assessment rating. Sackett and Dreher (1982) pointed out that two different evaluation processes are commonly used in the assessment center. In one method referred to as the within exercise rating technique, candidates are rated on each dimension upon the completion of each exercise. In the second method, referred to by Sackett and Dreher (1982) as the American Telephone and Telegraph Company (AT&T) model, "evaluation is postponed until the completion of all exercises at which time the assessors share their observations and then rate the candidates on a series of dimensions" (p. 402). Because the within exercise dimension rating technique provides ratings on each dimension for each exercise this method has become the focus of attention of recent studies examining the underlying dimensionality of assessment center ratings.

Sackett and Dreher (1982), for example, analyzed data from three different assessment centers all of which used the within exercise model. In each sample they found little convergence of dimension ratings across exercises relative to the convergence of dimension ratings within an exercise. Similar results have been found by Neidig, Martin, and Yates (1979) and Archambeau (1979). Further, Sackett and Dreher (1982) using factor analysis to examine the underlying dimensionality of ratings obtained from multiple exercises, found clear evidence for exercise factors in each of the three assessment centers rather than dimension factors.

As noted earlier, different assessment center methods may

force raters to organize and process assessment center data in different ways. The three studies cited above have all used only the within exercise rating technique. This method may force the assessor to process information in terms of exercises and therefore rate in terms of exercises. A different set of results may have been obtained had a different methodology for evaluating and integrating the assessment center data been used. Consequently, the purpose of the present study was to examine the effects of two different methods of evaluating and integrating assessment center data on the dimensionality and convergent and discriminant validity of the assessment center ratings. The two methodologies examined were the within exercise rating technique and a modified version of the AT&T model. Specifically, it was hypothesized that differences in convergent and discriminant validity and the factor structure of ratings would be obtained contingent upon the way the assessment center method forced the assessor to process the information.

Method

Participants

Candidates (N=90) were assessed for department manager positions in a large east coast retail organization. Forty-five candidates were assessed using the within exercise model, and 45 candidates were assessed using the modified AT&T model. Each candidate was rated by four assessors resulting in 180 ratings for each of the two models.

Procedure

Exercises and Dimensions. Assessors observed candidates in three exercises and assessed them on six dimensions. The exercises were an In-Basket and two role plays; while the dimensions were: problem analysis; problem solution; leadership; people sensitivity; sense of urgency; and communication. In one role play exercise, the "staffer" role play (SRP), candidates assumed the role of department manager and were required to determine and resolve the problems of their assistant. In the second role play exercise, the "customer" role play (CRP), candidates were confronted with an irate customer and were again required to determine and resolve the problems presented. All six dimensions were rated in each of the exercises.

Manipulation. Candidate ratings were obtained from six assessment teams each comprised of four store managers. Assessors were randomly assigned to the teams. Each of the six assessment teams rated 15 candidates in each of the three exercises. Three teams used the AT&T model in making evaluations, and three teams used the within exercise model. The assessment teams were ramdomly assigned to the two experimental conditions. The only difference in the experimental conditions was at the point in the evaluation meeting where the dimension ratings for each exercise were obtained.

The following procedure was used to obtain assessor ratings on the dimensions with the AT&T model. First, all assessors in the team observed and recorded information on the candidates in each of the three exercises. Second, the recorded information for

a candidate relevant to a specific dimension from all three exercises was placed on an overhead. For example, information on the leadership dimension from the CRP, SRP, and IB was placed on the overhead for discussion. Third, based on this information, each assessor privately made an overall dimension rating of the candidate's performance on that dimension (e.g., an overall rating on the leadership dimension). Fourth, if needed, these ratings were shared and discussed until a consensus rating was obtained on the dimension. Finally, assessors then privately rated the candidate's performance on the dimension for each of the three exercises. For example, each assessor made separate leadership ratings for the CRP, the SRP, and the IB. This procedure was followed for each dimension until all dimensions had been evaluated for a candidate. The dimension ratings on all three exercises obtained in this final step were subjected to analyses.

In the within exercise condition, after observing an exercise, each assessor privately rated the candidate's performance on all six dimensions. This procedure was followed for each of the three exercises. These dimension ratings on the three exercises were subjected to analyses.

Results

Several analyses were conducted to determine whether the within exercise model and the AT&T model showed differences in convergent and discriminant validity and differences in the dimensionality of assessment center ratings. Campbell and Fiske (1959) proposed a systematic experimental procedure for analyzing

convergent and discriminant validities. Convergent validity is demonstrated when correlations among the same dimensions measured by different exercises (monotrait-heteromethod correlations) are significant and large enough to have practical value. These mean correlations for the two assessment center models are presented in the upper portion of Table 1. All correlations were significant for both methods. However, the AT&T model correlations are comparably higher than the within exercise model correlations indicating greater convergent validity among dimensions for the AT&T model.

According to Campbell and Fiske (1959), discriminant validity may be demonstrated in a number of ways. First, monotrait-heteromethod correlations should be larger than the correlations among ratings of different dimensions measured in different exercises (heterotrait-heteromethod correlations). As indicated in Table 1, the mean heterotrait-heteromethod correlations for both models are lower than the grand mean correlations for the monotrait-heteromethod correlations suggesting some evidence for discriminant validity. Second, the monotrait-heteromethod correlations should be higher than correlations among different dimensions measured by the same exercise (heterotrait-monomethod correlations). The heterotrait-monomethod correlations are presented in the bottom portion of Table 1. These correlations for both the AT&T model and the within exercise model are higher than the monotrait-heteromethod correlations indicating poor discriminant validity. However, it should be noted that the discriminant

validity for the within exercise model is poorer than for the AT&T model.

A further technique used to assess convergent and discriminant validity is with analysis of variance (Kavanaugh, MacKinney, & Wolins, 1971). This method is based on the correlations in the multitrait-multimethod matrix. The sources of variation in this ANOVA model are: variations due to Assessor (convergent validity); variations due to the Assessor by Dimension interaction (discriminant validity); variations due to the Assessor by Exercise interaction (source or exercise variation); and Error. The values calculated for these sources are presented in Table 2. The appropriate values for the cross model comparisons are the intraclass correlation coefficients (ICC) presented in the final column of Table 2.

Comparisons of the ICC values between the two models suggests conclusions similar to those discussed above. The ICC values for the Assessor sources suggest somewhat stronger convergent validity for the AT&T model. Further, the Assessor by Dimension ICC's suggest relatively low discriminant validity for both models, but higher discriminant validity for the AT&T model. Finally, the Assessor by Exercise ICC values suggest stronger source variance due to exercises for the within exercise model than for the AT&T model.

The dimensionality of the data from the two models was examined with principal axis factor analysis using varimax rotation. Since there were six dimensions and three exercises, six and three factor solutions were hypothesized as potentially

meaningful solutions. This approach to examining the dimensionality of the ratings was the same as that taken by Sackett and Dreher (1982). In both six factor solutions, the last three factors were very weak, with few salient loadings. Consequently, only the three factor solutions for the two models are presented in Tables 3 and 4.

As can be seen by examining Tables 3 and 4, the factor structures for these two models are clearly different. For the within exercise model (Table 3) very clear exercise factors are present with no complex loadings (i.e., dimensions loading on more than one factor). Although there is evidence of exercise factors in the factor structure of the AT&T model (Table 4), there are many more complex loadings present suggesting that the assessors are not making as clear a distinction between assessee beahvior in the different exercises.

Discussion

Differences in the convergent and discriminat validities and the dimensionality of the assessment center ratings was observed between the two models. Although the two conditions differed in only one aspect (i.e., the nature of the integration of the assessment center data), this variation produced a perceptible difference in the structure of assessment center ratings. An explanation for these obtained differences is that the different assessment center integration sessions forced the rater to process and organize the assessment center data in different ways.

This explanation is suggested by the pattern of obtained

results in the study. The within exercise model forced the assessor to organize the assessment center information by exercise. Contrarily, the AT&T model required the assessor to think about the assessment center information somewhat less in terms of exercises, and more in terms of dimensions. Given these differences between the two methods it is not surprising that: 1) the AT&T model showed somewhat greater convergent validity for the dimensions than the within exercise method; 2) the AT&T method showed somewhat less source variance due to exercises than the within exercise model; and 3) the data from the within exercise model resulted in much cleaner exercise factors than the AT&T model.

The integration session is only one way in which assessment centers vary. Assessment centers may vary in a number of other aspects such as the number and type of exercises used, the number of dimensions evaluated, the length and content of assessor training, and methods for recording and observing behavior. Such variations in assessment center procedures may result in even more profound differences in the types of data obtained from the assessment center. Such differences may in some cases affect the types of decisions and conclusions which are drawn about candidates. Future research should consider the rater as a source of variation in assessment center ratings. Assessment center methods which affect assessors' cognitive processes and assumptions which assessors bring to the assessment situation should be evaluated for their effect on assessment center data.

References

Archambeau, D. J. (1979). Relationships among skill ratings assigned in an assessment center. Journal of Assessment Center Technology, 2, 7-20.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 36, 81-105.

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.

Kavanaugh, M. J., MacKinney, A. C., & Wolins, L. (1971). Issues in managerial performance: Multitrait-multimethod analysis of ratings. Psychological Bulletin, 75, 34-49.

Landy, F. J. & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

Neidig, R. D., Martin, J. C., & Yates, R. E. (1979). The contribution of exercise skill ratings to final assessment center evaluations. Journal of Assessment Center Technology, 2, 21-23.

Rush, M. C., Thomas, J. C., & Lord, R. G. (1977). Implicit leadership theory: A potential threat to the internal validity of leader behavior questionnaires. Organizational Behavior and Human Performance, 20, 93-110.

Sackett, P. R. & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. Journal of Applied Psychology, 67, 401-410.

Shack, M. (1983). The implications of cognitive and interpersonal processes for assessment centers. Unpublished doctoral dissertation, Bowling Green State University, Bowling Green, OH.

Table 1

Dimension and Exercise Correlations for the AT&T Model and the Within Exercise Model

	AT&T Model		Within Exercise Model	
	Mean r	SD	Mean r	SD
Dimension (Monotrait-Heteromethod Correlations)				
Problem Analysis	.48	.19	.41	.06
Problem Solution	.50	.10	.30	.10
Leadership	.59	.10	.43	.07
People Sensitivity	.57	.11	.35	.07
Sense of Urgency	.57	.09	.32	.12
Communication	.53	.15	.40	.13
Grand Mean and SD	.54	.15	.37	.11
Heterotrait-Heteromethod Correlations				
Grand Mean and SD	.44	.13	.31	.09
Exercise (Heterotrait-Monomethod Correlations)				
Staffer Role Play	.73	.06	.63	.13
Customer Role Play	.66	.05	.69	.09
In-Basket	.55	.07	.71	.10
Grand Mean and SD	.65	.10	.68	.11

Note. Due to the listwise deletion of missing data, N=169 for the AT&T model and N=178 for the Within Exercise Model.

BEST COPY AVAILABLE

Table 2

ANOVA Results for Multitrait-Multimethod Matrices

AT&T Model

Source	df	MS	F Value	Variance Component	Intraclass Correlation
Assessor (A)	168	9.80	37.70*	.53	.67
A x Dimension	840	.55	2.12*	.10	.27
A x Exercise	336	1.48	5.68*	.20	.44
Error	1680	.26		.26	

Within Exercise Model

Source	df	MS	F Value	Variance Component	Intraclass Correlation
Assessor (A)	177	8.26	30.44*	.44	.62
A x Dimension	885	.44	1.61*	.06	.17
A x Exercise	354	2.47	9.11*	.37	.57
Error	1770	.27		.27	

Note. * $p < .01$.

Table 3

Rotated Factor Pattern for the Within Exercise Model.

	Within Exercise Model			
			Factors	
Dimensions	Exercises	I	II	III
Problem Analysis	Staffer Role Play			.79
Problem Solution	Staffer Role Play			.80
Leadership	Staffer Role Play			.85
People Sensitivity	Staffer Role Play			.48
Sense of Urgency	Staffer Role Play			.77
Communication	Staffer Role Play			.67
Problem Analysis	Customer Role Play		.79	
Problem Solution	Customer Role Play		.84	
Leadership	Customer Role Play		.86	
People Sensitivity	Customer Role Play		.64	
Sense of Urgency	Customer Role Play		.82	
Communication	Customer Role Play		.78	
Problem Analysis	In-Basket	.85		
Problem Solution	In-Basket	.91		
Leadership	In-Basket	.86		
People Sensitivity	In-Basket	.61		
Sense of Urgency	In-Basket	.83		
Communication	In-Basket	.83		

Note. Only factor loadings greater than .35 are presented.

Table 4

Rotated Factor Pattern for the AT&T Model.

AT&T Model				
		Factors		
Dimensions	Exercises	I	II	III
Problem Analysis	Staffer Role Play	.71	.38	
Problem Solution	Staffer Role Play	.82	.38	
Leadership	Staffer Role Play	.87		
People Sensitivity	Staffer Role Play	.58	.49	
Sense of Urgency	Staffer Role Play	.69	.41	
Communication	Staffer Role Play	.71	.38	
Problem Analysis	Customer Role Play	.44	.63	
Problem Solution	Customer Role Play		.76	
Leadership	Customer Role Play	.61	.43	
People Sensitivity	Customer Role Play		.77	
Sense of Urgency	Customer Role Play	.35	.71	
Communication	Customer Role Play	.50	.57	
Problem Analysis	In-Basket			.68
Problem Solution	In-Basket			.75
Leadership	In-Basket	.35		.63
People Sensitivity	In-Basket			.65
Sense of Urgency	In-Basket			.78
Communication	In-Basket			.64

Note. Only factor loadings greater than .35 are presented.